

Digitale Philologie: Korpuslinguistische Analyse in Antiken Texten

Stelios Chronopoulos

Seminar für Griechische und Lateinische Philologie der Universität Freiburg

8. November 2018

Inhalte

Textkorpus und -korpora

Definitionen und Herstellung

Digitale Werkzeuge für die Analyse von Textkorpora

Konkordanzer und Frequenzliste

Signifikanztests

Definitionen für Textkorpora

- ▶ Textkorpora in der Klassischen Philologie:
 - ▶ Texte von einem bestimmten Autor
 - ▶ Texte von einer bestimmten Gattung
 - ▶ Texte in einer bestimmten Dialekt
 - ▶ Texte von einer bestimmten Epoche
- ▶ Textkorpora in Linguistik
 - ▶ verschiedene Kriterien: Gattung, Epoche, ein- oder mehrsprachige Korpora, schriftliche Texte oder Transkriptionen von mündlichen Texten
 - ▶ ausschlaggebend: die Größe – die Korpora sollen von Maschinen „gelesen“ und analysiert werden

Die Herstellung von Textkorpora

- ▶ Textkorpora werden hergestellt: sie sind das Ergebnis einer Forschungstätigkeit
- ▶ Herstellung/Auswahl von Textkorpora aufgrund von wissenschaftlicher Fragestellung
- ▶ Textkorpora versehen mit:
 - ▶ Metadata: Information über den Text
 - ▶ Text-Markup: Information zu Elementen des Textes
 - ▶ linguistische Annotationen (ein spezieller Fall von Text-Markup): linguistische Information zu Elementen des Textes

Konkrdanzer: Hauptsächliche Funktion

- ▶ Suche in einem Textkorpus nach bestimmten Sequenzen von Elementen (string)
- ▶ Darstellung/Visualisierung der Suchergebnisse

Frequenzliste: Hauptsächliche Funktion

- ▶ Elemente zählen in einem Textkorpus (es gibt x):
 - ▶ Wörter (tokens)
 - ▶ Wortformen (types)
- ▶ die Häufigkeit von Elementen zählen in einem Textkorpus (x ist häufig)
 - ▶ Wort/Wörter im Korpus
 - ▶ Wortform/Wort Ratio
- ▶ die Häufigkeit von Kookkurrenzen von Elementen zählen (x+y ist häufig)
 - ▶ Kookkurrenz von Wörtern (collocation)
 - ▶ Kookkurrenz von Wörtern und Konstruktionen
 - ▶ Kookkurrenz von Konstruktionen
 - ▶ Kookkurrenz von Wörtern und Stellungen im Text

Ziel der Signifikanztests

- ▶ Eliminieren die Wahrscheinlichkeit, ein Ergebnis zufällig zu sein
- ▶ Was ist ein zufälliges Ergebnis?

Signifikanztests für Schlüsselwörter und Kookkurenzen

- ▶ Schlüsselwörter: Vergleich zu einem Referenzkorpus
- ▶ Kookkurenzen: Mutual Information Tests und Z-score Tests